Poly-omic statistical methods describe cyanobacterial metabolic adaptation to fluctuating environments

Extended Abstract

Supreeta Vijayakumar
Department of Computer Science and Information Systems, Teesside University
Southfield Road, Tees Valley
Middlesbrough, United Kingdom TS1 3BX
s.vijayakumar@tees.ac.uk

Claudio Angione
Department of Computer Science and Information Systems, Teesside University
Southfield Road, Tees Valley
Middlesbrough, United Kingdom TS1 3BX
c.angione@tees.ac.uk

ABSTRACT

In this work, a genome-scale metabolic model of *Synechococcus* sp. PCC 7002 which utilizes flux balance analysis across multiple layers is analyzed to observe flux response between 23 growth conditions. This is achieved by setting reactions involved in biomass accumulation and energy production as objectives for bi-level linear optimization, thus serving to improve the characterization of mechanisms underlying these processes in photoautotrophic microalgae. Additionally, the incorporation of statistical techniques such as *k*-means clustering and principal component analysis (PCA) contribute to reducing dimensionality and inferring latent patterns.

CCS CONCEPTS

• **Applied computing** → **Systems biology**; *Biological networks*; *Bioinformatics*;

KEYWORDS

Synechococcus, phototrophic growth, multi-objective optimization, flux balance analysis, machine learning

ACM Reference format:
Supreeta Vijayakumar and Claudio Angione. 2017. Poly-omic statistical methods describe cyanobacterial metabolic adaptation to fluctuating environments. In *Proceedings of 9th International Workshop on Bio-Design Automation, Pittsburgh, Pennsylvania USA, August 2017 (IWBDA'17),* 2 pages.
https://doi.org/10.475/123_4

1 INTRODUCTION

Metabolic modelling can provide an intuitive way of monitoring the amount of change in essential biological pathways; e.g. reactions involved in cellular growth and repair, energy production, transport, etc. Genome-scale metabolic models (GSMMs) can be used to improve prediction of phenotypic outcomes through supplementing linear constraints for conducting flux balance analysis (FBA) with external data from multi-omic studies. However, this undertaking is often challenging owing to the persistent challenges of integrating multiple disparate data types [9]. The application of statistics and data mining can help to inform the inter-connectivity of these datasets when they are combined to glean meaningful information. *Synechococcus* sp. PCC 7002 is a fast-growing cyanobacterium which flourishes in both freshwater and marine environments, owing to its ability to tolerate high light intensity and a wide range of salinities. Harnessing the properties of cyanobacteria has become an imperative goal in recent years, owing to its potential for producing renewable biofuels [4]. Here, we evaluate the efficiency of *Synechococcus* sp. PCC 7002 as a chassis for biofuel production over various growth conditions, with the aim of optimizing biomass and energy production during photosynthesis.

IWBDA'17, August 2017, Pittsburgh, Pennsylvania USA

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

2 METHODS

We begin by calculating flux under phototrophic growth in a model of *Synechococcus* sp. PCC 7002 [4] using multi-omics flux balance analysis (FBA) [1] to obtain condition-specific flux profiles. Transcriptomic data was acquired in the form of RNA sequencing reads from a series of studies previously conducted by Ludwig and Bryant [6–8]. These data are loaded into the model using METRADE to map gene expression data to a space where each metabolic profile is associated with a different growth condition [2]. Normalised flux distributions are calculated using three pairs of objectives: (i) Biomass and ATP maintenance (ii) Biomass and Photosystem I, and (iii) Biomass and Photosystem II. The structure for bi-level linear optimization is given in (1), where FBA is carried out using the COBRA Toolbox in MATLAB. The f and g Boolean vectors weight objectives for FBA, while the $v^{\min}$ and $v^{\max}$ vectors represent lower- and upper-limits for flux rates. The product of the stoichiometric matrix of all metabolites and reactions (S) and the vector of flux rates for all reactions (v) is 0 as rates of metabolite consumption and production remain constant.

$$\begin{aligned} &\max && g^{\mathsf{T}} v \\ &\text{such that} && \max\, f^{\mathsf{T}} v, \quad Sv = 0, \\ &&& v^{\min}\varphi(\Theta) \le v \le v^{\max}\varphi(\Theta), \end{aligned} \tag{1}$$

In (2), Θ represents a vector of gene set expression values of the reactions associated with the fluxes in v, which are mapped to a coefficient for the lower- and upper-limits of the corresponding reaction by function φ, defined as:

$$\varphi(\Theta) = [1 + \gamma\,|log(\Theta)|]^{\mathrm{sgn}(\Theta-1)}\,. \tag{2}$$

PCA was conducted using the FactoMineR package in R [5] (pictured in 2) and produces a scree plot of percentage contributions to variance in the first five dimensions, as well an individual factor map where growth conditions are described by reaction fluxes

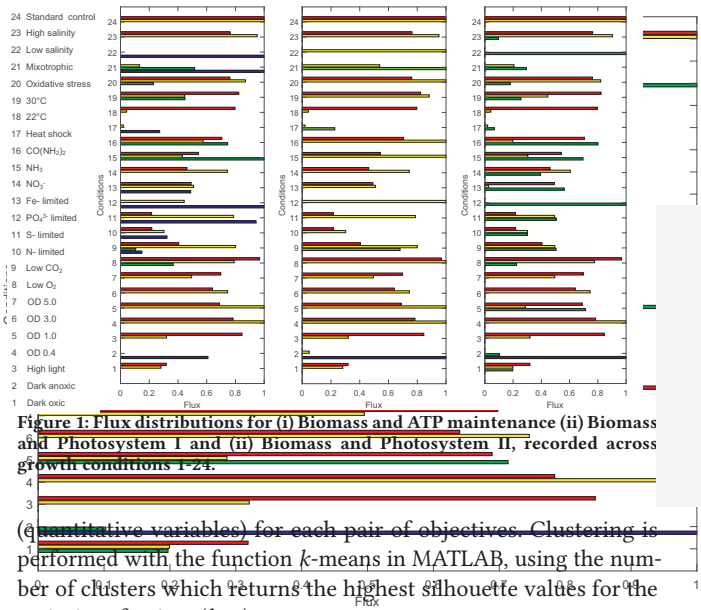

Figure 1: Flux distributions for (i) Biomass and ATP maintenance (ii) Biomass and Photosystem I and (ii) Biomass and Photosystem II, recorded across growth conditions 1-24.

(quantitative variables) for each pair of objectives. Clustering is performed with the function *k*-means in MATLAB, using the number of clusters which returns the highest silhouette values for the majority of points (*k*=6).

3 RESULTS AND DISCUSSION

By prioritising different pairs of objectives during FBA, the mechanisms underlying each pathway become more evident. When ATP maintenance is set as the secondary objective (Fig 1), the highest fluxes occur in heat shock and growth-limiting conditions, illustrating the importance of this reaction in maintaining cellular function when growth rate or energy transfer through the photosystems is low. Absence of light and oxygen are shown to lead to a significant decrease in growth, owing to lower generation of ATP and NADPH without photoautotrophic growth; nutrient (particularly phosphate) limitation also results in low biomass flux, compared to the control. The tolerance of *Synechococcus* sp. PCC 7002 for high light intensity is evident from high flux for all three objective pairs through the biomass pathway. For the high salinity condition, fluxes through biomass and photosystem I are high for all objective pairs, but flux is only maintained in the low salinity condition for the reaction set as the secondary objective *g*. When set as objective *g*, flux through photosystem II for the low salinity condition is much higher than the high salinity condition.

Principal components analysis (PCA) was carried out across the flux distributions generated for all objective pairs to identify the conditions and/or reactions responsible for the most variance in the datasets. Fig 2 displays a scree plot with the percentage of variance explained by the first dimensions and also an individual factor map, which displays the principal component scores of 24 individuals (simulated conditions) described by 742 variables (reactions) on the first two dimensions. For all three objective pairs, more than 70% of the variance can be explained by just two dimensions i.e. two linear combinations of all fluxes (2). Low oxygen, high light intensity, high salt, and lower temperature give the highest scores for the first dimension; these conditions also yield the highest fluxes in 1 and are in concordance with experimental findings [3, 7, 10]. For the second dimension, the highest score is given by low salt, mixotrophic and phosphate-limitation conditions for the ATP objective. *k*-means

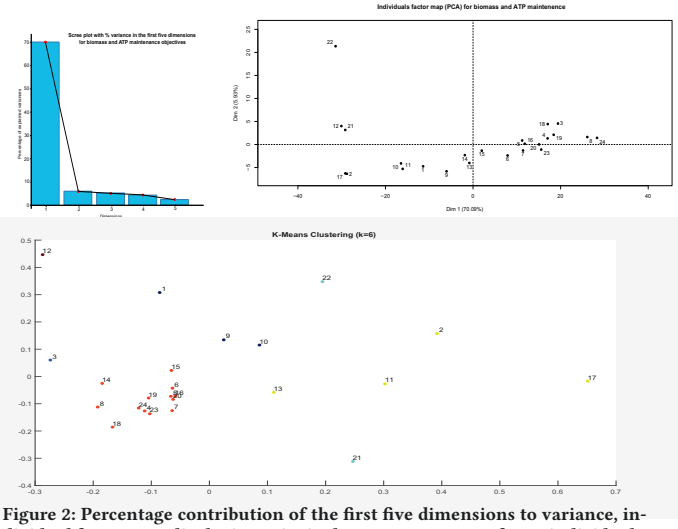

Figure 2: Percentage contribution of the first five dimensions to variance, individual factor map displaying principal component scores for 24 individuals (conditions) described by 742 variables (fluxes) on the first two principal components and *k*-means clustering performed with six clusters.

also reflects the flux distributions in showing clear differentiation between conditions. In accordance with their biomass fluxes, high light intensity and phosphate limitation are isolated from all other conditions. The grouping of mixotrophic and low salt conditions is indicative of their lack of flux through the biomass reaction. Other conditions which are detrimental for growth form two separate clusters- 1, 9, 10 and 2, 11, 13, 17. In the first group of conditions, it can be noted that some growth is maintained through biomass synthesis, whereas in the second, there is higher flux through the photoexcitation reactions for photosynthesis, potentially with reliance on the ATP maintenance pathway to drive this process.

4 CONCLUSIONS

The use of a condition-specific metabolic model, which incorporates gene expression data and assesses multiple objectives, allows for prediction of significant metabolic patterns and phenotypic outcomes arising as a result of adaptation to fluctuating environmental conditions. In addition to this, statistical techniques such as PCA and clustering introduce another layer of analysis for uncovering latent patterns by re-organizing data on the basis of shared characteristics, therefore providing further insight into the maintenance of metabolic efficiency during phototrophic growth.

REFERENCES

[1] Claudio Angione, Max Conway, and Pietro Lió. 2016. *BMC Bioinformatics* 17, 4 (2016), 257.
[2] Claudio Angione and Pietro Lió. 2015. *Sci. Rep.* 5 (2015), 15147.
[3] Hans C Bernstein, Ryan S McClure, Eric A Hill, Lye Meng Markillie, William B Chrisler, Margie F Romine, Jason E McDermott, Matthew C Posewitz, Donald A Bryant, Allan E Konopka, et al. 2016. *mBio* 7, 4 (2016), e00949–16.
[4] John I Hendry, Charulata B Prasannan, Aditi Joshi, Santanu Dasgupta, and Pramod P Wangikar. 2016. *Bioresour. Technol.* 213 (2016), 190–197.
[5] Sébastien Lê, Julie Josse, and François Husson. 2008. *J. Stat. Softw.* 25, 1 (2008), 1–18.
[6] Marcus Ludwig and Donald A Bryant. 2011. *Front. Microbiol.* 2 (2011), 41.
[7] Marcus Ludwig and Donald A Bryant. 2012. *Front. Microbiol.* 3 (2012), 354.
[8] Marcus Ludwig and Donald A Bryant. 2012. *Front. Microbiol.* 3 (2012), 145.
[9] Supreeta Vijayakumar, Max Conway, Pietro Lió, and Claudio Angione. 2017. *Briefings in Bioinformatics* (2017).
[10] Qian Xiong, Jie Feng, Si-ting Li, Gui-ying Zhang, Zhi-xian Qiao, Zhuo Chen, Ying Wu, Yan Lin, Tao Li, Feng Ge, et al. 2015. *Mol Cell Proteomics* 14, 4 (2015), 1038–1053.